# Scene Experts: Specializing in 3D Gaussian Splatting with Adaptive Decomposition

**Xiaowen Fu, Yang Zhang*, Yuhan Tang, Huazhong Zhang, Tianxing Zhao, Yuhang Guo, Yu Huang, Jinbao Wang**

Shenzhen University
2410673014@mails.szu.edu.cn, yangzhang@szu.edu.cn

## Abstract

Anchor-based 3D Gaussian Splatting (GS), exemplified by Scaffold-GS, achieves remarkable storage efficiency through a hybrid explicit-implicit representation. However, their reliance on a single, monolithic network to decode anchor features imposes a severe bottleneck on model capacity, often resulting in blurred details and view-dependent artifacts in complex scenes. To break this bottleneck, we introduce the concept of Scene Experts: a strategy that decomposes the task of modeling a complex scene across a collection of specialized sub-models. To realize the paradigm, we propose MoE-GS. Our approach designs the decoder as a Sparsely-Gated Mixture of Experts (MoE), which dramatically increases the model's total capacity while maintaining comparable inference cost via sparse activation. To effectively train this high-capacity model, we propose two key innovations: (1) A progressive curriculum learning strategy that first trains all experts on a robust baseline before encouraging them to specialize on different scene components. (2) A novel opacity-aware regularization that penalizes inactive neural Gaussians, ensuring the expanded capacity is efficiently used. Extensive experiments demonstrate that MoE-GS substantially outperforms state-of-the-art methods on diverse benchmarks, significantly improving reconstruction fidelity while requiring a smaller or comparable Gaussian model size.

## 1 Introduction

Scene reconstruction is a pivotal task in 3D vision. Recently, 3D Gaussian Splatting (3DGS) (Kerbl et al. 2023) has pioneered a new paradigm for the task, introducing an explicit representation. Specifically, 3DGS models a scene as a collection of anisotropic 3D Gaussians. By employing a differentiable rasterization pipeline, 3DGS fully leverages the parallel architecture of GPUs to achieve high-fidelity, real-time rendering. Despite its breakthrough, 3DGS faces a curse of scale (Ali et al. 2025). It represents the scene using millions of Gaussians, each of which requires storing explicit parameters for position, covariance, opacity and color. This reliance on a massive set of primitives results in a substantial storage footprint. To mitigate this, a solution is to adopt an anchor-based framework that introduces the implicit representation into the fully explicit 3DGS (Lu et al. 2024; Chen et al. 2024; Wang et al. 2024; Zhang et al. 2025). This strategy embeds scene information within a compact neural network, thereby achieving a significant reduction in model size while ensuring rendering quality and efficiency.

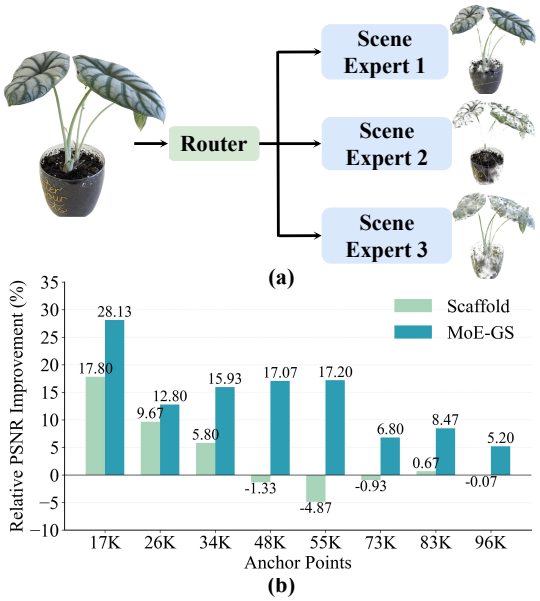


Figure 1: Scene Experts for adaptive decomposition and performance comparison. We propose Scene Experts, a paradigm where a router adaptively decomposes an object or scene into distinct components, each handled by a scene expert (a). This approach effectively resolves the capacity bottleneck of prior methods. As shown in (b), our MoE-GS consistently and significantly outperforms Scaffold-GS in relative PSNR improvement over 3DGS, especially in anchor-dense scenes, where Scaffold-GS's performance degrades. From left to right: playroom, room, kitchen, Quebec, Pompidou, garden, stump, and bicycle. Our method and Scaffold-GS produce a comparable number of anchor points across these scenes. Details are provided in the appendix.

However, while these anchor-based methods succeed in model compression, their core design introduces a new bottleneck: **the reliance on a single, monolithic MLP to de-**

*Corresponding author: Yang Zhang.

**code features for the entire scene.** As shown in Figure 1(b), when the number of anchors or the scene complexity increases, the expressive power of this single MLP becomes a performance bottleneck. It struggles to capture the diverse and intricate geometric and appearance details across the scene, imposing a ceiling on the rendering quality.

To overcome the performance ceiling, we introduce the Scene Experts, a set of specialized sub-models designed to replace the monolithic decoder. We hypothesize that by decomposing the global modeling task into smaller, more focused sub-problems, each expert can specialize in a distinct aspect of the scene. This division of labor allows for a significant increase in the model's total representational power. The success of applying Mixture of Experts (MoE) (Du et al. 2022; Riquelme et al. 2021) in NeRF (Zhenxing and Xu 2022; Cong et al. 2023) has demonstrated the power of specialized modeling. Guided by these, we propose MoE-GS, a framework that realizes the Scene Experts paradigm within anchor-based 3DGS. Our approach designs the decoder as a sparsely activated MoE layer. A lightweight router learns to dynamically assign different anchors to specialized scene experts in an end-to-end manner, shown in Figure 1(a). However, we find that direct sparse training leads to premature specialization, causing experts to fall into local optima because each is only exposed to a subset of anchors early on. To mitigate this, we propose a curriculum learning strategy, where the training begins with the Dense MoE, allowing all experts to learn a robust baseline. Subsequently, we transition to the Sparse MoE, where experts are encouraged to specialize. To ensure the balanced utilization of this expanded capacity, we also introduce a load balancing loss.

Furthermore, we identify a secondary bottleneck in Scaffold-GS: a large portion of the neural Gaussians have negative opacities, shown in Table 3, meaning they are inactive and do not contribute to rendering. This keeps the number of active Gaussians far below its theoretical potential. To unleash the full potential of our high-capacity MoE model, we introduce a novel opacity-aware regularization loss. This loss encourages decoded opacities to be positive, thereby increasing the activation rate of neural Gaussians and fully leveraging the model's expanded capacity. Extensive experiments across various datasets demonstrate that our method achieves high-quality novel view synthesis with a similar or even smaller model size, outperforming current state-of-the-art (SOTA) anchor-based methods.

In summary, our contributions are summarized below:

- To address the capacity bottleneck of the monolithic MLP in anchor-based 3DGS, we propose MoE-GS, a novel MoE framework guided by a curriculum learning strategy. It enables an end-to-end scene decomposition, allowing the Scene Experts to model different parts, which enhances the model's capacity and expressiveness.
- We introduce a novel opacity regularization loss that significantly increases the activation rate of neural Gaussians, ensuring the expanded capacity is fully utilized.
- Extensive experiments on benchmarks show that MoE-GS achieves SOTA rendering quality, advancing the frontier of efficient and high-fidelity scene reconstruction.

# 2 Related Work

## 2.1 Anchor-based 3D Gaussian Splatting

Scaffold-GS (Lu et al. 2024) addresses the significant redundancy in 3DGS (Kerbl et al. 2023) by introducing a compact anchor-based representation, where a lightweight MLP decodes Gaussian attributes from anchor features. This reliance on a single monolithic MLP, however, imposes a severe capacity bottleneck for complex scenes. While follow-up works enhance compression through entropy coding (Chen et al. 2024; Wang et al. 2024) or feature expressiveness with second-order anchors (Zhang et al. 2025), they still inherit the bottleneck of the single MLP.

## 2.2 Mixture of Experts

MoE has become a dominant architecture for scaling up networks. The idea is to replace a dense network with multiple specialized experts, governed by a simple routing mechanism. The router dynamically selects a subset of experts for each input, enabling conditional computation. The approach is prominently applied in language models (Shazeer et al. 2017), and is further refined by works like GShard (Lepikhin et al. 2020) and Switch Transformers (Fedus, Zoph, and Shazeer 2022) through load balance and simplified routing strategies. Recently, MoE has been explored for scene representation within NeRF (Mildenhall et al. 2021; Barron et al. 2021; Pumarola et al. 2021; Martin-Brualla et al. 2021). While prior works have applied MoE to NeRF for large-scale reconstruction (Zhenxing and Xu 2022) or multi-scene generalization (Cong et al. 2023), they are coupled with implicit representation and are not directly applicable for 3DGS. In contrast, MoE-GS is the first to propose Scene Experts for anchor-based 3DGS methods, which addresses the capacity limitation of them by decomposing the scene for specialized modeling.

# 3 Method

We introduce MoE-GS, a novel anchor-based 3DGS method designed for superior rendering quality. We propose the Scene Experts paradigm with an opacity-aware regularization loss. Figure 2 shows the overview of our method.

## 3.1 Preliminary

**Scaffold-GS.** Scaffold-GS (Lu et al. 2024) introduces a structured anchor-based scene representation method, using a set of anchor points $\{v_i\}$ to dynamically generate neural Gaussians in local regions. These anchors are initialized from the SfM (Fisher et al. 2021) points and distributed in a sparse voxel grid. Each anchor $v$ is equipped with a position $x_v$, a feature vector $f_v \in \mathbb{R}^D$, a scaling factor $l_v \in \mathbb{R}^{k\times 6}$, and offsets $O_v \in \mathbb{R}^{k\times 3}$, which are all learnable parameters. Unlike 3DGS, Scaffold-GS dynamically derives $k$ neural Gaussians from each anchor on the fly during rendering. The attributes of these Gaussians, including the scale $s_{v,k}$, rotation $q_{v,k}$, color $c_{v,k}$, and opacity $\sigma_{v,k}$, are predicted by a set of lightweight, shared MLPs. The inputs to these MLPs are the anchor's feature $f_v$, the relative distance $\delta_{vc}$, and the direction $d_{vc}$ between the camera and the anchor. Specifically, the opacities of $k$ neural Gaussians are predicted by

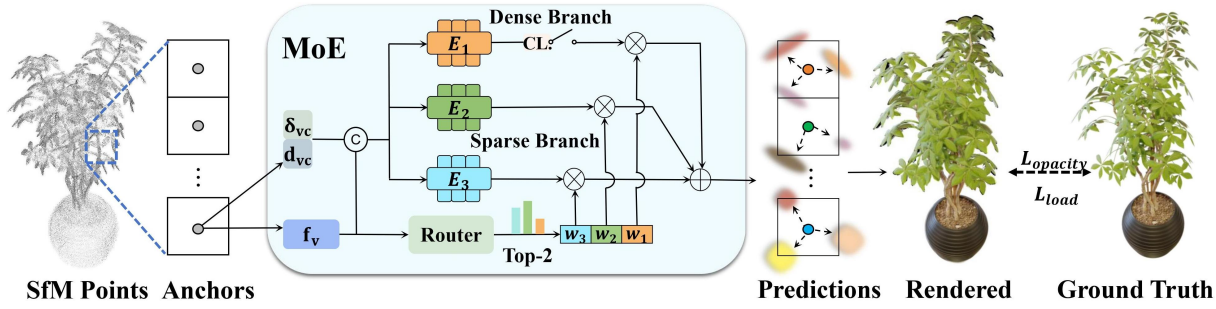


Figure 2: Overview of MoE-GS. Initialized from SfM points, each anchor is equipped with a learnable feature vector $f_v$, which is fed into a router network that outputs gating weights for the scene experts $E_1, E_2, E_3$. We employ a dual-strategy curriculum learning (CL) strategy. During a warmup phase, we use the dense MoE setting. Then we switch to the sparse Top-k MoE, allowing experts to specialize on different levels of the scene's appearance. $f_v$ is concatenated with the distance to camera $\delta_{vc}$ and the view direction $d_{vs}$, which is processed by the selected experts. The final output is a weighted sum of their predictions. To fully leverage the expanded model capacity, we incorporate a load balancing loss $L_{load}$ and an opacity-aware regularization loss $L_{opacity}$ during training. We use 3 experts and the Top-2 strategy as an example.

the MLP $F_\sigma$:

$$\{\sigma_{v,0}, \ldots, \sigma_{v,k-1}\} = F_\sigma(f_v, \delta_{vc}, d_{vc}).$$

Other attributes, such as color, scale, and rotation, are predicted in parallel by MLPs $(F_c, F_s, F_q)$. The positions for $k$ Gaussians are computed by adding the corresponding learnable offsets to the anchor's position:

$$\{\mu_0, ..., \mu_{k-1}\} = x_v + \{O_0, ..., O_{k-1}\} * l_v.$$

### 3.2 MoE-Enhanced Representation Learning

**Expert and Router Architecture.** For each Gaussian attribute like opacity, covariance, and color, we employ a separate and independent MoE layer. This design choice grants maximal flexibility, allowing the model to learn distinct specializations for different properties. Each MoE layer consists of $N$ scene experts, each sharing the same architecture as the original MLP, and a lightweight router network $R$. The router is to predict an optimal assignment of anchors to the experts. To make the routing decision structure and appearance aware, the router takes the anchor's learned feature vector $f_v$ as input. The router outputs a vector of logits $\mathbf{s} \in \mathbb{R}^N$. These are normalized using a softmax function to produce the gating weights $\mathbf{g}(f_v)$, representing the router's confidence in assigning the anchor to each expert:

$$\mathbf{g}(f_v) = \text{softmax}(R(f_v)).$$

Some method (Zhenxing and Xu 2022) uses point positions as routing inputs; however, our experiments (included in the appendix) show that incorporating anchor positions does not lead to performance improvement.

**Dual-Strategy Curriculum Learning.** The experts tend to specialize prematurely during the early stages of training as they are initially exposed to only a limited portion of the scene. This prevents them from learning a robust baseline, causing the final reconstruction quality to lag behind that of a dense configuration. To address this, we introduce a curriculum learning approach that leverages two different MoE strategies for effective training.

**Phase 1: dense MoE warm-up.** To balance effective initial training for all experts with computational efficiency, we introduce a warm-up period that precedes the densification. For the first $T_{\text{warmup}}$ training iterations, we employ a dense MoE formulation. The output for a given attribute is computed as a weighted average of all expert outputs, using the gating weights $\mathbf{g}(f_v)$ from the router:

$$y(f_v) = \sum_{i=1}^{N} g_i(f_v) E_i(f_v).$$

Here, $E_i$ is the $i$-th expert. The dense activation forces all experts to participate in the learning process from beginning, ensuring all experts develop a meaningful representation.

**Phase 2: sparse Top-k specialization.** After the initial warm-up period, we switch to a sparse Top-k MoE strategy. For each anchor, only the Top-k experts with the highest gating probabilities are activated. Let $\mathcal{I} = \text{TopKIndices}(\mathbf{g}(f_v), k)$ be the set of indices for the Top-k experts. The gating weights for these selected experts are re-normalized to sum to one:

$$w_i(f_v) = \frac{g_i(f_v)}{\sum_{j \in \mathcal{I}} g_j(f_v)}, \quad \forall i \in \mathcal{I}.$$

The final output is then a weighted sum of only the activated experts' outputs. The transition to sparse activation encourages experts to specialize in different subsets of the scene, while significantly reducing the computational load.

**Load Balancing.** To prevent the router from favoring only a few experts, we incorporate an auxiliary load balancing loss $L_{load}$ (Shazeer et al. 2017; Dai et al. 2024; Omi, Sen, and Farhadi 2025). This loss encourages the router to distribute anchors evenly across all experts, as shown in the results in Figure 3. Let $\mathcal{B}$ be the set of anchors in a training batch. We define $\pi_i$ as the fraction of anchors assigned to the expert $E_i$ and $P_i$ as the average routing probability for this expert across the batch. The loss is then:

$$L_{load}^S = N \sum_{i=1}^{N} \pi_i \cdot P_i.$$

Here, $\pi_i = \frac{1}{|\mathcal{B}|}\sum_{f_v \in \mathcal{B}} \mathbb{I}(i \in \mathcal{I})$, $P_i = \frac{1}{|\mathcal{B}|}\sum_{f_v \in \mathcal{B}} g_i(f_v)$.

Furthermore, we address load balancing during our dense warm-up phase. Standard load balancing loss like $L^S_{load}$ is designed for sparse routing and is not applicable here. Therefore, we propose $L^D_{load}$, tailored for dense weighted assignments. The core objective is to ensure that each expert receives a comparable contribution over a training batch. For a batch of $B$ input samples, the router assigns a continuous weight vector $\mathbf{g}(f_v)$ to each input $f_v$, where $\sum_{i=1}^{N} g_i(f_v) = 1$. We define the average contribution for each expert over the batch as: $C_i = \frac{1}{B}\sum_{\boldsymbol{v} \in \boldsymbol{\mathcal{B}}} g_i(f_v)$. To achieve balance, we want the contribution to be distributed nearly evenly among all experts. The average routing probability is mathematically equivalent to its contribution, so our proposed loss for dense MoE $L^D_{load}$ is formulated as:

$$L^D_{load} = N \sum_{i=1}^{N} (C_i)^2.$$

This loss is proportional to the variance of the expert contribution distribution, effectively preventing model collapse where only a subset of experts are consistently utilized.

### 3.3 Loss Design

Our model is optimized through a composite loss function that balances rendering fidelity with model compactness and efficiency. The rendering loss $\mathcal{L}_{\text{render}}$ and the scale regularization $\mathcal{L}_{\text{scale}}$ are adopted from Scaffold-GS. Additionally, we incorporate the MoE load-balancing loss $\mathcal{L}_{\text{load}} = L^S_{load} + L^D_{load}$, as detailed in Section 3.2, to promote an even workload distribution among the experts.

We observe that a number of predicted neural Gaussians have negative opacity values during training. These inactive or dead Gaussians do not contribute meaningfully to the rendered scene, indicating an under-utilization of model capacity. To mitigate this, we introduce a regularization term that promotes more active participation from all Gaussians. Instead of aggressively pushing all opacities to 1, our goal is to discourage Gaussians from becoming inactive. We achieve this by penalizing opacity $\sigma$ that falls below a certain threshold $\tau$ (e.g., 0). The loss is formulated as a hinge loss:

$$\mathcal{L}_{\text{opacity}} = \mathbb{E}_{\mathcal{G}}[max(0, \tau - \sigma)],$$

where the expectation is taken over all Gaussians $\mathcal{G}$. The term encourages that most Gaussians could remain alive and actively contribute to the scene, shown in Table 3.

The final objective function combines these components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{render}} + \lambda_{\text{opacity}}\mathcal{L}_{\text{opacity}} + \lambda_{\text{scale}}\mathcal{L}_{\text{scale}} + \lambda_{\text{load}}\mathcal{L}_{\text{load}},$$

where $\lambda_{\text{opacity}}$, $\lambda_{\text{scale}}$, and $\lambda_{\text{load}}$ are hyperparameters that balance the influence of each regularization term.

## 4 Experiment

### 4.1 Experimental Setup

**Dataset.** We conduct our experiments using the same datasets as Scaffold-GS (Lu et al. 2024). We select 7 scenes from the MipNeRF360 (Barron et al. 2022) dataset, 2 scenes from Tanks&Temples (Knapitsch et al. 2017), and 2 scenes from DeepBlending (Hedman et al. 2018). These scenes cover a wide range of environments, from bounded indoor environments to unbounded outdoor settings. Additionally, we include 6 scenes from the BungeeNeRF (Xiangli et al. 2022) dataset to assess our model's effectiveness in complex, large-scale outdoor scenarios. We follow the official training/testing splits and image resolutions from 3DGS.

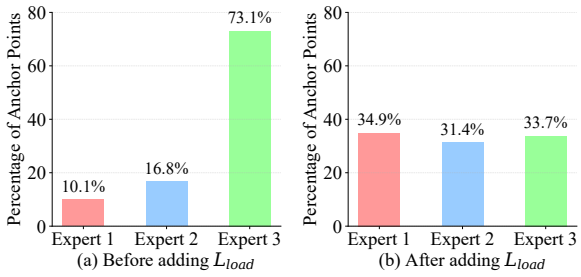


Figure 3: Ablation study on $L_{load}$. It shows the percentage of anchors allocated to each expert relative to the total (using the 3-expert with Top-2 as an example) before (a) and after (b) incorporating $L_{load}$. Without $L_{load}$, the router tends to assign most anchor points to a single expert. Introducing $L_{load}$ significantly balances the expert workload. This example uses the covariance-predicting expert on the Quebec scene from BungeeNeRF. Similar trends are observed with different numbers of experts and with $L^D_{load}$.

**Methods and Metrics.** Following the evaluation strategy of SOGS (Zhang et al. 2025), we benchmark MoE-GS against anchor-based 3DGS methods, as this line of work is most relevant to our contributions. Our primary baselines are the original 3DGS and Scaffold-GS. We also compare ours with SAGS (Ververas et al. 2024) and Octree-GS (Ren et al. 2025), which can be integrated with Scaffold-GS or share similar principles. We intentionally exclude methods like HAC (Chen et al. 2024) and Context-GS (Wang et al. 2024), as they focus on the orthogonal problem of model compression, whereas our work targets a different goal. Due to the lack of a public implementation for SOGS, a fair retraining on our device and settings is not feasible. Nevertheless, our method also surpasses the rendering metrics reported in the original paper. We evaluate the methods utilizing the commonly used PSNR, SSIM (Wang et al. 2004), and LPIPS (Zhang et al. 2018) metrics. Besides, we report model storage requirements in megabytes (MB).

**Implementation Details.** For each Gaussian attribute, we employ a separate set of experts and a corresponding router, which is a simple linear layer. Each expert retains the original MLP architecture from Scaffold-GS, which consists of two linear layers with a ReLU activation. To evaluate our method, we conduct experiments using a Top-k routing strategy for k=1 and k=2. We explored various configurations by varying the number of experts from 2 to 5. Furthermore, a dense model was included in our evaluation to serve as a performance upper bound. We set $T_{warmup}$ to the first 1500 iterations, after which the densification begins. Regarding loss design, the weights $\lambda_{\text{opacity}}$ and $\lambda_{\text{load}}$ are both set to

| Scene | MipNeRF360 | | | | DeepBlending | | | | Tanks&Temples | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Metrics | PSNR ↑ | SSIM ↑ | LPIPS ↓ | size(MB) | PSNR ↑ | SSIM ↑ | LPIPS ↓ | size(MB) | PSNR ↑ | SSIM ↑ | LPIPS ↓ | size(MB) |
| 3DGS | 29.47 | 0.882 | 0.131 | 614.0 | 29.70 | 0.899 | 0.267 | 619.2 | 23.74 | 0.853 | 0.170 | 373.1 |
| Scaffold-GS | 29.61 | 0.881 | 0.136 | 166.4 | 30.19 | 0.902 | 0.272 | 56.0 | 24.02 | 0.854 | 0.174 | 77.0 |
| SAGS | 30.20 | 0.880 | 0.138 | 319.3 | 29.92 | 0.899 | 0.284 | 162.8 | 24.12 | 0.841 | 0.207 | 186.1 |
| Octree-GS | 29.59 | 0.881 | 0.134 | 176.7 | 30.27 | 0.901 | **0.264** | 64.3 | 24.47 | 0.865 | **0.153** | 170.5 |
| MoE-GS | **30.23** | **0.888** | **0.125** | **158.0** | **30.64** | **0.905** | **0.264** | **55.4** | **24.61** | **0.866** | **0.153** | **75.7** |

Table 1: Quantitative comparison of novel view synthesis. MoE-GS achieves the best rendering quality while slightly reducing model size. It uses 5 experts with a Top-2 routing strategy across all datasets, which our ablation study shows strikes the best balance between performance and efficiency. We compare only with the most relevant anchor-based methods, excluding HAC and Context-GS, as they focus more on model compression.

| | PSNR ↑ | SSIM ↑ | LPIPS ↓ | size(MB) |
|---|---|---|---|---|
| 3DGS | 27.84 | 0.914 | 0.100 | 1561.1 |
| Scaffold-GS | 27.94 | 0.912 | 0.109 | 171.6 |
| SAGS | 28.52 | 0.920 | 0.120 | 423.2 |
| Octree-GS | 28.23 | 0.922 | 0.090 | 333.4 |
| MoE-GS | **28.64** | **0.925** | **0.089** | **161.6** |

Table 2: Quantitative comparison on large-scale scenes from the dataset BungeeNeRF.

| | Activation Ratio (%) ↑ | | |
|---|---|---|---|
| Scene | MipNeRF360 | DeepBlending | Tanks&Temples |
| Scaffold-GS | 36.9 | 41.0 | 34.6 |
| + $L_{opacity}$ | **63.1** | **70.5** | **66.8** |

Table 3: Ablation study on $L_{opacity}$. The Activation Ratio (%) is defined as the percentage of Gaussians with a positive opacity. The results show that including $L_{opacity}$ significantly boosts activation across datasets, indicating effective utilization of model capacity.

0.01, while other loss hyperparameters follow Scaffold-GS. The threshold $\tau$ for opacity regularization is set to 0, thereby encouraging the activation of more neural Gaussians. Other key hyperparameters remain consistent with the settings in the original 3DGS and Scaffold-GS. All experiments are conducted on NVIDIA A6000 GPUs using PyTorch 2.7.

## 4.2 Comparision with the State-of-the-Art

Table 1 shows that our MoE-GS consistently outperforms all methods across all datasets and metrics. On the BungeeNeRF, Table 2 shows that MoE-GS achieves the most significant gains over Scaffold-GS, with a PSNR improvement of over 0.7 dB. Our MoE-GS framework leverages five scene experts with a Top-2 routing, a configuration shown in ablation studies to offer the best balance between quality and efficiency. This design overcomes the capacity bottleneck of a monolithic model, allowing specialized experts to effectively model diverse visual aspects, which is critical for reconstructing large and complex scenes. Our method further reduces the Gaussian model size of Scaffold-GS. This reduction is attributed to our opacity-aware regularization, which efficiently leverages the expanded implicit modeling capabilities and thereby mitigates the need for a large number of anchors. Although more expert networks are introduced, they are lightweight enough that their storage overhead is negligible compared to the Gaussian model.

In Figure 4, we provide a visual comparison between all methods. With the reduced model size, our method demonstrates superior capabilities in novel view synthesis, yielding renderings with fewer artifacts and sharper details.

## 4.3 Scene Decomposition

Figure 5 shows the specialization of our experts. We separately render the neural Gaussians predicted by the anchors routed to each expert in the setting with $N = 3$ and the Top-2 strategy for the clear visualization. It reveals that the experts implicitly decompose the scene into different levels, which validates the efficacy of our MoE framework. Notably, the crucial insight here is not the interpretability of the decomposition, but its effectiveness in enabling specialization similar to a divide-and-conquer strategy. More results are shown in the appendix.

## 4.4 Ablation Studies

To validate the effectiveness of our MoE framework and the proposed loss, we conduct ablation studies by removing these components from our full model. Our full model is configured with five experts with Top-2, reported in Table 4.

**Mixture of Experts.** Ablating the MoE module results in a significant degradation in performance. This result strongly underscores the effectiveness of our core innovation, validating that specialized experts can better model the complex appearance of a scene than a single monolithic network. Notably, even in this ablated configuration, our model still outperforms Scaffold-GS across all reported metrics. This remaining advantage is primarily attributed to the contribution of our opacity-aware regularization.

**Curriculum Learning.** The removal of this strategy (CL) leads to a discernible drop in performance across all metrics. This result highlights the importance of the approach in starting from the robust baselines, helping the model avoid suboptimal local minima.

**Expert Balancing.** Ablating $L_{load}$ results in a consistent performance drop. This demonstrates its importance in maintaining a balanced expert contribution, which is crucial for effectively leveraging the expanded capacity of the MoE. As shown in Figure 3, without an explicit balancing strategy, the router tends to assign a majority of anchor points to a

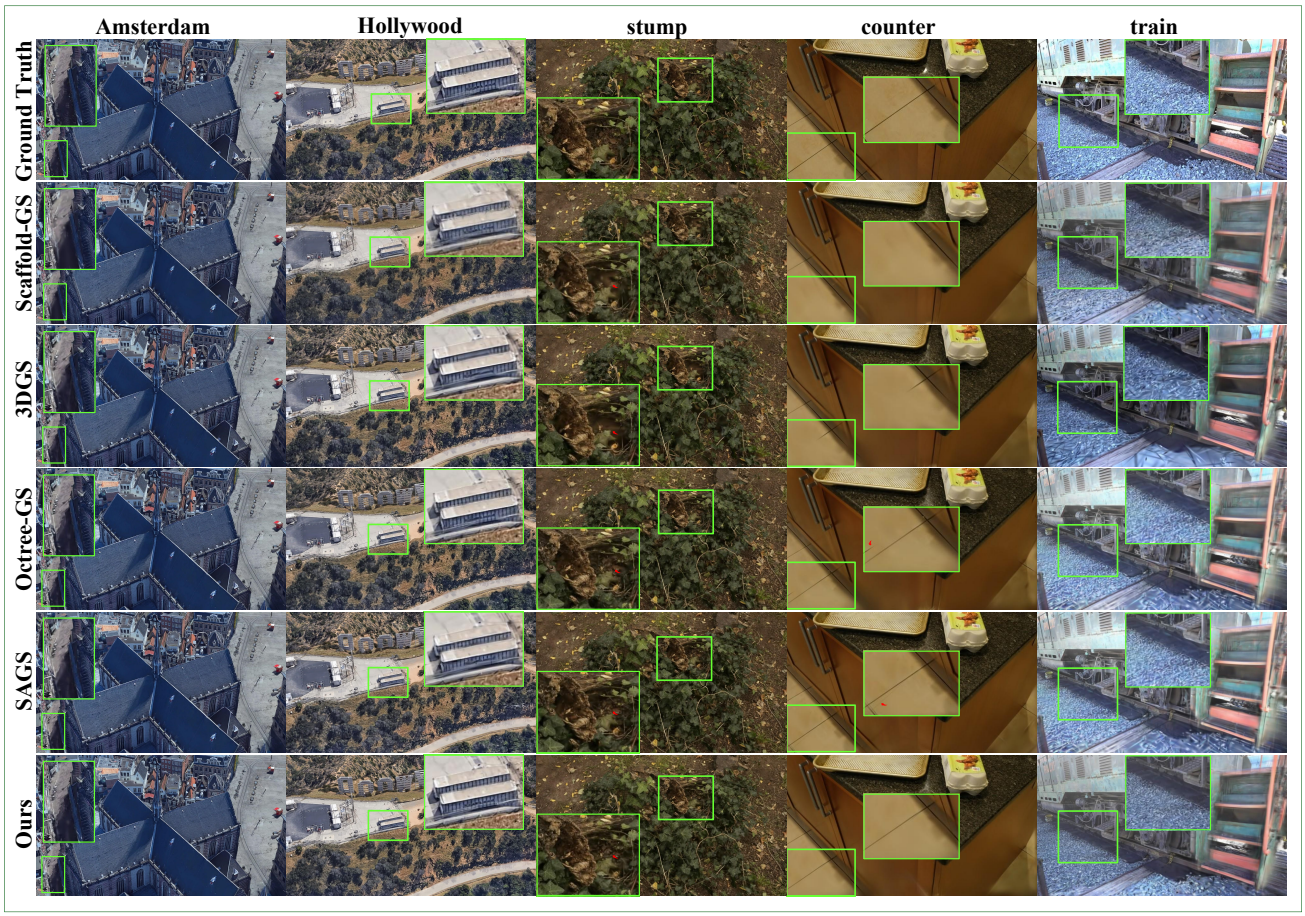


Figure 4: Qualitative comparisons of all methods. MoE-GS achieves renderings with much fewer artifacts and fine-grained details. Zoom in to view details.

| Scene | DeepBlending | | | | Tanks&Temples | | | |
|---|---|---|---|---|---|---|---|---|
| Ablation & Metrics | PSNR ↑ | SSIM ↑ | LPIPS ↓ | size(MB) | PSNR ↑ | SSIM ↑ | LPIPS ↓ | size(MB) |
| w/o MoE | 30.37 | 0.902 | 0.272 | 53.2 | 24.30 | 0.858 | 0.167 | 74.6 |
| w/o CL | 30.47 | 0.903 | 0.265 | 55.9 | 24.42 | 0.864 | 0.156 | 76.1 |
| w/o $L_{load}$ | 30.53 | 0.904 | 0.264 | 56.0 | 24.47 | 0.865 | 0.154 | 77.7 |
| w/o $L_{opacity}$ | 30.52 | 0.903 | 0.271 | 58.6 | 24.46 | 0.864 | 0.159 | 78.8 |
| MoE-GS | **30.64** | **0.905** | **0.264** | **55.4** | **24.61** | **0.866** | **0.153** | **75.7** |

Table 4: Ablation study on the components of MoE-GS. The framework uses 5 experts with Top-2 across the datasets.

single expert. After incorporating the $L_{load}$ loss, the expert load becomes significantly more balanced.

**Opacity-aware Regularization.** The removal of $L_{opacity}$ not only leads to a noticeable decline in quality but also results in a larger model size. This impact powerfully confirms its effectiveness, directly contributing to both higher performance and greater model efficiency. As shown in Table 3, Scaffold-GS exhibits a low activation ratio. However, with the addition of our $L_{opacity}$, this ratio increases significantly, which demonstrates the effectiveness of our proposed loss.

## 4.5 Analysis on MoE Hyperparameters

We conduct a detailed analysis of two key hyperparameters within our MoE framework: the total number of experts $N$ and the routing strategy. All experiments were performed on the DeepBlending, with results visualized in Figure 6. We use the total training time as a practical proxy for the computational cost of each configuration.

**Effect of the Number of Experts ($N$).** As shown in Figure 6, increasing $N$ generally yields higher PSNR across all routing strategies, including activating all experts (Dense). This result validates our initial motivation: expanding the model's capacity by adding more experts is an effective strategy for improving reconstruction performance. The setting with five experts consistently demonstrates superior performance compared to the others. Therefore, we select $N = 5$ as the configuration for our model.

**Effect of Top-k Routing.** The results indicate the high ef-


Figure 5: Per-expert output visualizations with 3 experts and Top-2. The Scene Expert 1 appears to handle certain structural edges and high-contrast areas. The Scene Expert 2 focuses on modeling broader, more continuous surfaces. The Scene Expert 3 captures the scene's foundational color by modeling color-consistent regions. More results are included in the appendix.

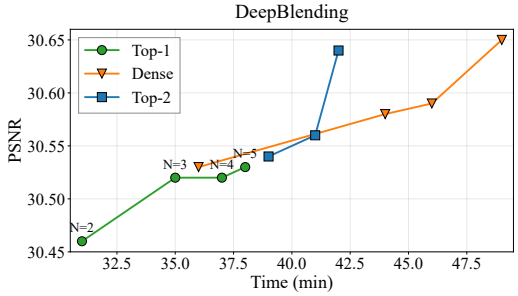


Figure 6: Analysis on MoE Hyperparameters. It shows that reconstruction quality scales with the number of experts. However, this presents a trade-off: the Top-1 strategy is maximally efficient but compromises quality, whereas a dense setup achieves superior quality at the highest computational overhead. We identify that a configuration of 5 experts with Top-2 provides the best trade-off between reconstruction fidelity and efficiency. Other datasets show similar trends.

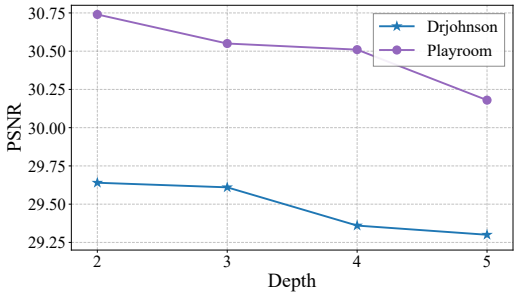


Figure 7: Analysis of network depth on rendering quality. We find that not all approaches to increasing model capacity yield performance gains. Specifically, simply increasing the number of linear layers in Scaffold-GS fails to improve rendering quality and even leads to performance degradation. Other datasets show similar trends.

ficiency of sparse routing. While the Dense strategy sets the performance upper bound, it does so at the highest computational expense. The Top-1 strategy is the fastest but yields comparatively lower performance. This limitation becomes more pronounced at $N = 4$ and $N = 5$. As the number of experts grows, the model learns to decompose the scene into more fine-grained visual components. In such cases, selecting a single expert is insufficient to integrate the diverse attributes. The Top-2 strategy (3 to 5 experts), however, strikes an optimal balance. It delivers performance nearly on par with the dense model while being significantly more efficient in training time. We adopt Top-2 in our final model. This approach allows for near-ensemble performance with high efficiency, demonstrating its key advantage.

### 4.6 Limitations of Naive Capacity Scaling

We investigate alternative strategies for expanding model capacity. While increasing network depth is an intuitive approach, our experiments reveal its limitations in this context. In Figure 7, when we progressively increased the MLP depth within Scaffold-GS, rendering quality did not improve but instead degraded. In contrast, our method provides a viable scaling path, as shown in Figure 6. This finding underscores the effectiveness of MoE-GS, which enhances the model's expressive capacity by decomposing the scene across multiple scene experts in an end-to-end manner.

## 5 Conclusion

We propose Scene Experts, a new paradigm for anchor-based 3DGS. We identify a performance bottleneck in these methods, like Scaffold-GS, where a single network struggles to effectively model anchor-dense scenes. Our approach overcomes this by employing the Scene Experts, each dedicated to learning different aspects of the scene's appearance. These experts are orchestrated by an end-to-end learned router within our framework, MoE-GS, which adaptively delegates tasks to the most suitable expert. The decomposition and specialization significantly enhance the model's total capacity and lead to superior rendering results. Furthermore, we address the issue of the large portion of inactive neural Gaussians by designing an opacity-aware regularization loss, ensuring the expanded model capacity is fully utilized. Extensive experiments demonstrate that MoE-GS achieves superior rendering quality with a similar or more compact model. By integrating a dynamic, high-capacity model, MoE-GS advances the frontier of hybrid explicit-implicit 3D scene reconstruction.

## Acknowledgements

The work was supported by National Natural Science Foundation of China (Grant No. 62176163), Shenzhen Higher Education Stable Support Program General Project (Grant No. 20231120175215001), Scientific Foundation for Youth Scholars of Shenzhen University, Science and Technology Foundation of Shenzhen (Grant No. JCYJ20210324094602007), National Natural Science Foundation of China (Grant No. 62320106007). and National Key R&D Program of China (Grant No. 2024YFF0618403).

## References

Ali, M. S.; Zhang, C.; Cagnazzo, M.; Valenzise, G.; Tartaglione, E.; and Bae, S.-H. 2025. Compression in 3d gaussian splatting: A survey of methods, trends, and future directions. *arXiv preprint arXiv:2502.19457*.

Barron, J. T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; and Srinivasan, P. P. 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5855–5864.

Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5470–5479.

Chen, Y.; Wu, Q.; Lin, W.; Harandi, M.; and Cai, J. 2024. Hac: Hash-grid assisted context for 3d gaussian splatting compression. In *European Conference on Computer Vision*, 422–438. Springer.

Cong, W.; Liang, H.; Wang, P.; Fan, Z.; Chen, T.; Varma, M.; Wang, Y.; and Wang, Z. 2023. Enhancing nerf akin to enhancing llms: Generalizable nerf transformer with mixture-of-view-experts. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3193–3204.

Dai, D.; Deng, C.; Zhao, C.; Xu, R.; Gao, H.; Chen, D.; Li, J.; Zeng, W.; Yu, X.; Wu, Y.; et al. 2024. Deepseekmoe: Towards ultimate expert specialization in mixture-of-experts language models. *arXiv preprint arXiv:2401.06066*.

Du, N.; Huang, Y.; Dai, A. M.; Tong, S.; Lepikhin, D.; Xu, Y.; Krikun, M.; Zhou, Y.; Yu, A. W.; Firat, O.; et al. 2022. Glam: Efficient scaling of language models with mixture-of-experts. In *International conference on machine learning*, 5547–5569. PMLR.

Fedus, W.; Zoph, B.; and Shazeer, N. 2022. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120): 1–39.

Fisher, A.; Cannizzaro, R.; Cochrane, M.; Nagahawatte, C.; and Palmer, J. L. 2021. ColMap: A memory-efficient occupancy grid mapping framework. *Robotics and Autonomous Systems*, 142: 103755.

Hedman, P.; Philip, J.; Price, T.; Frahm, J.-M.; Drettakis, G.; and Brostow, G. 2018. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics (ToG)*, 37(6): 1–15.

Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4): 139–1.

Knapitsch, A.; Park, J.; Zhou, Q.-Y.; and Koltun, V. 2017. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36(4): 1–13.

Lepikhin, D.; Lee, H.; Xu, Y.; Chen, D.; Firat, O.; Huang, Y.; Krikun, M.; Shazeer, N.; and Chen, Z. 2020. Gshard: Scaling giant models with conditional computation and automatic sharding. *arXiv preprint arXiv:2006.16668*.

Lu, T.; Yu, M.; Xu, L.; Xiangli, Y.; Wang, L.; Lin, D.; and Dai, B. 2024. Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20654–20664.

Martin-Brualla, R.; Radwan, N.; Sajjadi, M. S.; Barron, J. T.; Dosovitskiy, A.; and Duckworth, D. 2021. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7210–7219.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Omi, N.; Sen, S.; and Farhadi, A. 2025. Load Balancing Mixture of Experts with Similarity Preserving Routers. *arXiv preprint arXiv:2506.14038*.

Pumarola, A.; Corona, E.; Pons-Moll, G.; and Moreno-Noguer, F. 2021. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10318–10327.

Ren, K.; Jiang, L.; Lu, T.; Yu, M.; Xu, L.; Ni, Z.; and Dai, B. 2025. Octree-GS: Towards Consistent Real-time Rendering with LOD-Structured 3D Gaussians. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–15.

Riquelme, C.; Puigcerver, J.; Mustafa, B.; Neumann, M.; Jenatton, R.; Susano Pinto, A.; Keysers, D.; and Houlsby, N. 2021. Scaling vision with sparse mixture of experts. *Advances in Neural Information Processing Systems*, 34: 8583–8595.

Shazeer, N.; Mirhoseini, A.; Maziarz, K.; Davis, A.; Le, Q.; Hinton, G.; and Dean, J. 2017. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*.

Ververas, E.; Potamias, R. A.; Song, J.; Deng, J.; and Zafeiriou, S. 2024. Sags: Structure-aware 3d gaussian splatting. In *European Conference on Computer Vision*, 221–238. Springer.

Wang, Y.; Li, Z.; Guo, L.; Yang, W.; Kot, A.; and Wen, B. 2024. Contextgs: Compact 3d gaussian splatting with anchor level context model. *Advances in neural information processing systems*, 37: 51532–51551.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Xiangli, Y.; Xu, L.; Pan, X.; Zhao, N.; Rao, A.; Theobalt, C.; Dai, B.; and Lin, D. 2022. Bungeenerf: Progressive neural radiance field for extreme multi-scale scene rendering. In *European conference on computer vision*, 106–122. Springer.

Zhang, J.; Zhan, F.; Shao, L.; and Lu, S. 2025. SOGS: Second-Order Anchor for Advanced 3D Gaussian Splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 11167–11176.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

Zhenxing, M.; and Xu, D. 2022. Switch-nerf: Learning scene decomposition with mixture of experts for large-scale neural radiance fields. In *The Eleventh International Conference on Learning Representations*.